Terrain classification using ToF sensors for the enhancement of agricultural machinery traversability

Francisco Yandun Narváez[a,*], Eduard Gregorio[b], Alexandre Escolà[b], Joan R. Rosell-Polo[b], Miguel Torres-Torriti[c], Fernando Auat Cheein[a]

[a] *Electronic Engineering Department, Universidad Técnica Federico Santa María, Valparaíso, Chile*
[b] *Research Group on AgroICT and Precision Agriculture, Department of Agricultural and Forest Engineering, Universitat de Lleida-Agrotecnio Center, Lleida, Spain*
[c] *School of Electrical Engineering, Pontificia Universidad Católica de Chile, Valparaíso, Chile*

Abstract

Ground properties influence various aspects of mobile machinery navigation including localization, mobility status or task execution. Excessive slipping, skidding or trapping situations can compromise the vehicle itself or other elements in the workspace. Thus, detecting the soil surface characteristics is an important issue for performing different activities in an efficient, safe and satisfactory manner. In agricultural applications, this point is specially important since activities such as seeding, fertilizing, or ploughing are carried on within off-road landscapes which contain a diversity of terrains that modify the navigation behaviour of the vehicle. Thus, the machinery requires a cognitive capability to understand the surrounding terrain type or its characteristics in order to take the proper guidance or control actions. This work is focused on the soil surface classification by implementing a visual system capable to distinguish between five usual types of off-road terrains. Computer vision and machine learning techniques are applied to characterize the texture and color of images acquired with a Microsoft Kinect V2 sensor. In a first stage, development tests showed that only infra-red and RGB streams are useful to obtain satisfactory accuracy rates (above 90%). The second stage included field trials with the sensor mounted

*Corresponding author. Tel.: +56 975209248
Email address: francisco.yandun.14@sansano.usm.cl (Francisco Yandun Narváez)

on a mobile robot driving through various agricultural landscapes. These scenarios did not present illumination restrictions nor ideal driving roads; hence, conditions can resemble real agricultural operations. In such circumstances, the proposed approach showed robustness and reliability, obtaining an average of 85.20% of successful classifications when tested along 17 trials within agricultural landscapes.

Keywords: Agricultural robotics, terrain classification, pattern recognition.

---

1. Introduction

Autonomous navigation within agricultural scenarios is a particularly challenging point since they are semi-structured environments composed of human workers, animals, obstacles and rough terrain that limit the machinery mobility and constraint its movement. In order to drive along a feasible, safe and efficient path, these vehicles must be capable of being aware of their surroundings for dealing with the constant changes of such elements within the workspace. In general, this capability is related with both object recognition (to interact with the environment or to extract information from it), and the vehicle-terrain interaction. The last point is specially challenging since the diversity of soil types present in the agricultural fields makes the scenes usually consisting of low-traction, deformable and steep-hill terrains, which can quickly degenerate the quality of the positioning and compromise the task execution. Therefore, control and path planning strategies can be executed based on the classification and characterization of the driving terrain or its surroundings (Iagnemma and Ward, 2009). In the same way, management of machinery resources (e.g., battery or fuel) can be improved, increasing the aggregated value of the agricultural activity (Michel, 2012; Xue et al., 2012). Furthermore, the integrity of the vehicle itself can be preserved by avoiding excessive slipping, skidding or even trapping conditions.

Interpretation and characterization of the terrain surface have been studied using dynamic and descriptive methodologies. Dynamic terramechanical ap-

proaches are performed based on known wheel-soil models which provide information about the tractive forces involved during the navigation (Al-Milli et al., 2010; Taheri et al., 2015). In addition, longitudinal and lateral slip effects of such terra-mechanical interaction can be measured (Botha and Els, 2015b,a). This knowledge, along with the kinematic characteristics of the robot, have proven to enhance the traversability of wheeled mobile robots, specially in Mars rovers (Ishigami et al., 2009; Brooks and Iagnemma, 2012). In these cases, the unavoidable limitation regarding the lack of a priori in-situ information about the terrain, requires the robot to drive through it first in order to obtain feedback (in relation to terrain interaction) from propioceptive sensors on board such as accelerometers or encoders (Brooks and Iagnemma, 2005; Ojeda et al., 2006). Thus, using additional sensors to explore the robot surroundings for predicting its navigation behaviour is not plausible, specially when the vehicle is driving through completely new scenes. Unlike these scenarios, agricultural landscapes do not present this drawbacks, since it is possible -and in some cases necessary- to use available a-priori information in order to anticipate the upcoming navigation behaviour of the vehicle.

A descriptive characterization of the terrain consists in representing the most relevant properties of the soil surface as ground planes, elevation maps or terrain classification. The first two are plausible alternatives; however, the continuously changing characteristics of the farming lands make these methodologies impractical in some situations (e.g., when vegetation has grown or the soil has been ploughed). Terrain classification, on the other hand, can be performed during a normal operation and would allow to know beforehand the upcoming terrain and discern whether a region is traversable or not, based on the latest information acquired by the sensors (Ho et al., 2013; Ball et al., 2016).

Various challenging points arise when using exteroceptive and propioceptive information for these purposes. For example, the ambient conditions of field operations (e.g., weather conditions or vibration) can often limit the measurement capabilities. The computational cost and processing capabilities have to guarantee practical applications. Another important point to concern is the to-

tal cost of the solution since agricultural applications like autonomous wheeled machinery are aimed to be commercially adopted by farmers. From this point of view, a trade-off between robustness and cost should be achieved to develop applications that would impact in the agricultural industry.

Given the previous context, this research work is focused on the use of a low cost sensor to obtain a descriptive interpretation of the terrain. A study and application of a classification system capable of determining the type of soil surface in front of a vehicle navigating on agricultural landscapes is proposed. Infra-red, color and depth streams are acquired with the sensor and used in a supervised classification scheme. It is shown that only infra-red information complemented with color is required to obtain high accuracy classification rates in real-time during field operations. The proposed system is tested using a mobile robot driving through a variety of agricultural terrains. It is noteworthy that this research work is not intended to determine terramechanic variables of the terrain, but to provide a complementary system that used together with those that employ wheel-ground models (Ishigami et al., 2009; Gao et al., 2013) could enhance the traversability assessment capabilities of an autonomous mobile robot.

This brief is organized as follows: Section 2 presents a review of the state of the art regarding terrain classification for diverse applications. Section 3 describes the hardware employed, as well as the methodology developed in this work. In Section 4, the experimental results, consisting of a validation and real condition tests, are shown. In Section 5 we provide the analysis and discussion of the results obtained, specially for the real condition tests. Finally, in Section 6 we present the conclusions of our work.

2. Related Work

Terrain characteristics influence directly in the navigation performance of wheeled mobile robots, specially in off-road scenarios like the agricultural landscapes (Prado et al., 2016). In this sense, it is reported the use of a variety of

sensors, models and processing algorithms to provide a descriptive interpretation of the soil surface. Propioceptive information obtained with inertial and vibration sensors have been used in various studies to distinguish between several types of terrain (Brooks and Iagnemma, 2012; Park et al., 2012). However, these approaches require the vehicle to drive through the terrain first to obtain a label, which make them impractical in various situations (e.g., driving over excessively muddy soil can result in a trapping situation). To overcome this drawback, exteroceptive sensors (or a combination with propioceptive sensors) have been employed with promising results. Specially, the robustness and working versatility in field conditions of light detection and ranging (LiDAR) or radar have made these devices commonly used (Reina et al., 2012; Fernandez, 2010). Furthermore, the first winner of the Defense Advanced Research Projects Agency (DARPA) challenge (Thrun et al., 2006) used several 2D LiDARs in a fusion scheme to classify terrain in front of the robot as occupied, free and unknown. The resulting map, along with the other systems of the robot, allowed it to autonomously drive more than 6 hours through the Mojave desert in the United States. Despite of the results reached with such range sensors, they have shortcomings that make them not suitable in certain cases that include presence of fog, excessive dust, smoke, or specular properties of the surfaces. Further, a "finer" classification of the terrain in various categories is difficult to obtain, as reported by Andújar et al. (2013).

Vision sensors have also been widely used in this context. From the color or spectral information provided by these devices, parameters such as textural or geometrical features can be obtained. For example, Ono et al. (2015) distinguished various classes of traversable or non traversable regions by employing grayscale images of a camera mounted on a space rover. For agricultural purposes, the same issue has also been addressed by using single color cameras or stereo-vision systems (Manduchi et al., 2005; Ball et al., 2016). Moreover, Zou et al. (2014) compared different approaches for terrain classification from visual information. The authors extracted different texture and color descriptors of images captured from moving robots in outdoor environments. The results

showed that using color and texture descriptors and learning algorithms with 2420 samples provided robustness to moderate changes in illumination. Precisely, the sensitivity of these approaches to changes in lightning makes them to require a high number of training images in order to cover the expected conditions, as reported by Angelova et al. (2007). In such work, a stereo camera and image processing techniques were used to classify 5 types of terrain based on texture descriptors of color images. The system used a supervised learning algorithm with approximately 3000 training samples, obtaining 76.4% of accuracy but no real time operation is reported.

Multi-sensor approaches can overcome the individual limitations of each device, providing robust and accurate solutions in real operation conditions. For instance, Häselich et al. (2013) used a fusion of color and three dimensional laser information to distinguish between 5 terrain classes. Various features from this data were calculated and subsequently applied to a probabilistic learning approach based on Markov Random Fields (MRF). The proposed system was tested in field with a mobile robot, obtaining accuracies greater than 95%, but real-time operation can not be performed when employing features from the two sensors. In addition, combination of LiDAR, stereovision, radar and thermography for detecting obstacles and traversable ground was studied by Reina et al. (2016). This work investigated diverse learning algorithms to combine data from these sensors in an off-line processing scheme

The total cost of the equipment is also an important point to take into account. Thus, approaches based on non-expensive sensors have also been studied with promising results in various agricultural applications (Rosell-Polo et al., 2015; Xia et al., 2015). Specifically, video gaming devices have been used as sensing systems for terrain classification, reporting promising results (Falola, 2012; Woods et al., 2015). Further, describing soil surface under 4 different tillage operations was studied in Marinello et al. (2015) by using a Kinect V1 sensor from Microsoft Corporation. Roughness features were obtained based on depth streams of a commercial structured-light camera. However, these type of sensors are intended to be used indoors, since their outdoors performance seri-

ously decrease. Recently, a new version of the Kinect device based on the time of flight principle was released, providing outdoors robustness for certain applications. A previous work of the authors of this research showed the usability of such sensor for terrain classification in conditions similar to real operations (Yandun et al., 2016). Results showed high accuracy ($> 80\%$), even with streams acquired in different weather conditions using a mobile setup.

Diverse algorithms and methodologies have been studied for processing data acquired with the previously discussed sensors. Image processing, probabilistic methods or supervised learning are reported to be employed in diverse works (Broten et al., 2012; Zou et al., 2014). However, recent studies have opted for self-learning approaches that reduced or removed the need of the training stage in learning algorithms (Otsu et al., 2016; Reina et al., 2012). These techniques usually consist in implementing two classifiers, one of them is trained with few samples or unsupervised learning. The second classifier uses the outcome of the first one to continuously learn and subsequently determine the corresponding labels. Thus, the versatility of the system is increased, providing adaptability to new scenes. Nevertheless, the resulting accuracy is prone to high error when the first classifier provides misslabelings. Furthermore, the availability of plenty a priori in-situ information (specially for agricultural applications) is an important point that solutions should take advantage.

3. Materials and Methods

In this work, we employed the second generation of the Kinect device (Microsoft Corporation, USA). The sensor provides infra-red (IR), color and depth streams at a maximum of 30 frames per second. In order to measure infra-red reflectance and estimate depth, it uses an IR camera with 512×424 pixel resolution. On the other hand, the color camera has a 1920×1080 pixel resolution. Infra-red measurements are based on active sensing, whereas depth is estimated based on a time of flight principle. It basically consists in measuring the amount of light received by synchronized detectors working in a complementary mode.

This configuration provides increased noise rejection and better accuracy when compared with its predecessor. In addition, the sensor has a built-in ambient light rejection that detects if a pixel is over saturated and resets the reflectance value measured by such pixel (Lau, 2013). These characteristics make the sensing device versatile and robust enough in a variety of outdoor applications. Specially, for terrain recognition, the authors studied the suitability of using this device to distinguish between several agricultural soil surfaces (Yandun et al., 2016).

3.1. Experimental Setup

We used two experimental scenarios to validate and test our system: static and real-dynamic operation. The static setup was employed to acquire the development dataset (i.e., training and testing data to establish and evaluate our approach). It consisted of placing the camera facing downwards to grab fixed depth, IR and color frames. For each terrain type studied in this brief, we collected 820 frames per class at different illumination conditions. Once the system was tested and validated, the sensor was mounted on the mobile robot Pioneer 3-AT from Mobile Robots Llc, in such a way that it was pointing downwards and forward, obtaining a view of the terrain in a range of 0.15 m to 0.9 m in front of the robot, as shown in Fig.1. This range is mainly affected by two aspects: i) the mounting of the sensor on the robot, and ii) the maximum range of the Kinect to obtain confident measures of depth (and therefore IR reflectance). Thus, according to the manufacturer, a maximum range of 4.5 m in front of the robot could be covered, with the proper setup of the sensor on the robot.

To test our approach in real conditions, the robot was manually driven at maximum linear and angular speeds of 0.5 m/s and 2.44 rad/s, respectively through various agricultural fields, acquiring and processing the IR and color streams at a frequency of 2 Hz.

Figure 1: Experimental setup employed to test our approach using a mobile robot driving through agricultural fields.

3.2. Architecture

The proposed terrain classification system employed image processing techniques as well as a supervised learning approach according to the layout depicted in Fig. 2. It allowed us to distinguish five types of agricultural soil surfaces: sand, grass, pavement, gravel and litterfall & straw- covered. These terrains can be cataloged surfaces of soil types from I to VIII according to the classification provided by Stolpe (2002) for the Chilean agricultural landscapes. Additionally, it was empirically found that traction of the robot changes when driving through each one. Thus, distinguishing between them can led to better energy management or avoiding dangerous hidden terrain topologies (it is not possible to know the geometry beneath litterfall or straw covered terrain) which would compromise the robot integrity (Prado et al., 2016). For field tests, we used only the IR and color streams provided by the sensor since it was experimentally found that depth information did not contribute to increase the classification accuracy, as will be shown later in Section 4. Therefore, only texture and predominant color

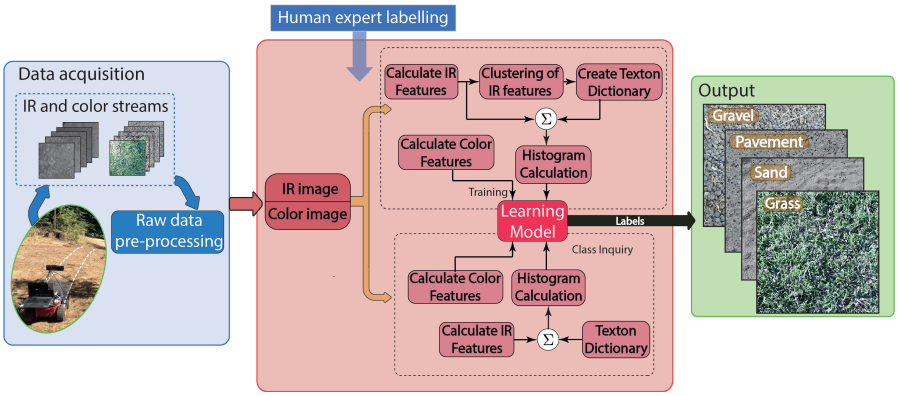


Figure 2: Block diagram of the terrain recognition system

of the terrains were analysed, setting aside its geometrical characteristics. Furthermore, our method highly relied on detecting the texture of the IR images, using color only as an additional feature to improve the classification accuracy.

The development and testing datasets were acquired at different illumination conditions and orientations, which caused various artifacts in the raw images, specially when excessive sunlight was present in the scene. To overcome this issue and to deal with the image boundary errors introduced by the sensor (Lachat et al., 2015), we cropped the raw images to retain only a central region of 300×300 pixels. In addition, we also replaced the missing information in specific pixels with the average intensity value of its neighbours. Subsequently, texture and color features were calculated to be used as training and testing inputs in a Support Vector Machine (SVM) classifier. In this work, we employed the *one vs one* multi-class approach of this algorithm due to its balanced training characteristic and applicability for various sizes of datasets, as shown by Hsu and Lin (2002).It was also chosen to assign one label per image, since the dimensions of the space covered by each one did not allow to obtain representative patches of different terrains. Thus, the output of our system provides a single label of the terrain type in front of the robot.

It is also noteworthy that data from all sensors were synchronously acquired using an application developed in C++ and Matlab (MathWorks, USA) in a

shared memory framework.

3.3. Terrain Classification

As stated before, the strategy employed to visually distinguish between terrain types is mainly focused on the textural characterization of the images provided by the sensor. As demonstrated in the authors' previous work (Yandun et al., 2016), and contrasting to the work of Angelova et al. (2007), using IR information can enhance the classification accuracy with a relative small train dataset. Further, including color or depth features along with the textural descriptors, could improve the capability of the method for detecting certain terrains. Thus, robustness in real operation conditions is also achieved, with reduced computational cost.

We employed the texture description methodology proposed by Varma and Zisserman (2002). It consists in building learning models based on local descriptors of each pixel and its statistical co-occurrence all over the image. Briefly, the most representative characteristics of the training images for each terrain class are first obtained using a clustering algorithm. Once grouped, such characteristics -called textons- represent the main descriptors of a single class. Subsequently, textons of all training images are collected in a single texton dictionary (TD). Once it is defined, a class model per training image in form of histogram can be obtained. It is built by labelling each pixel descriptor with the position of the closest texton in the dictionary. These histograms are the feature vectors used to train a supervised classifier. In the case of the query images, the previously built TD and the same methodology to obtain the histograms is applied in order to create the testing features.

Following this concept, various methods to obtain the local descriptors have been proposed, such as: filter responses (Varma and Zisserman, 2005), Local Binary Patterns (LBP) (Alvarez and Vanrell, 2012), Self-Invariant Feature Transform (SIFT) (Xu et al., 2012), among others. However, we employed the patch-based concept presented by Varma et al. (2009) since its results are comparable to those that apply a combination of various higher-dimensional

features (e.g., SIFT and its variants), as reported by Zhang et al. (2007). Thus, the approach presented in this work arises from the idea that textures can be considered as realizations of Markov Random Fields (MRF), as described by Li (2009). Formally, given a rectangular region of interest (ROI) S, and a set of m random variables defined in S which can take values $f = \{f_1, f_2, ..., f_m\}$ (e.g., pixel intensities), for a MRF it can be written:

$$p(f_i|f_{S-\{i\}}) = p(f_i|f_{\mathcal{N}_i}) \tag{1}$$

where f_i is the value of pixel i, $f_{S-\{i\}}$ represents the values of all pixels in S except i, and $f_{\mathcal{N}_i}$ stands for the set of pixel values in the $N \times N$ pixel neighbourhood (excluding the pixel i). Thus, the local description of a single pixel can be represented by the raw intensities of an $N \times N$ square neighbourhood of that point.

Thus, using this texture description methodology, we employed a 5×5 neighbourhood to obtain the local descriptors (i.e., a 25^{th} dimensional vector). Then, these vectors were clustered by using the standard K-means algorithm with $K = 10$, obtaining a total of 50 elements in the TD (5 classes $\times$ 10 centres). It is noteworthy that $K = 10$ was the value that yielded better classification accuracy after several trials in a validation stage. Finally, the feature space is created by 50-dimensional vectors that represent the frequency of occurrence of each TD element in the image.

In parallel to the previous processing, color information was employed to create an additional small feature vector. To this aim, the original RGB image was first converted to the -Comission Internationale de l'Éclairage (CIE) 1976-L*a*b* colorspace (Schwiegerling, 2004). The L* and b* components were left out for two reasons: i) to attenuate the effects of the lightning conditions, and ii) it was found in the development stage that they did not contribute to distinguish between classes. Thus, only the mean and variance values of the a* component were employed. In addition, the proportion of green in the image was calculated as: $\frac{G}{R+G+B}$, where R, G and B are the sum of intensity

Figure 3: Several snapshots of the robot driving through the experimental locations. Illumination varied from cloudy to clear sky conditions.

values per pixel for the R, G and B channels, respectively. When using depth streams in the validation tests, depth feature vectors consisted of four statistical measurements of roughness: mean, root mean square, skewness and kurtosis, as calculated by Marinello et al. (2015).

Finally, these descriptors (either depth or color) were aggregated to the texture feature vectors to form the input training and testing data in a multi-class Support Vector Machine classifier.

4. Experimental Results

Experiments for acquiring the development dataset and subsequently run the trials with the robot were performed in three locations: i) the Botanical Garden of Viña del Mar, located at -33.048093^o latitude and -71.500135^o longitude; ii) the Sporting Club of Valparaíso, located at -33.024568^o latitude and -71.532815^o longitude; and iii) the Technical University Federico Santa María (UTFSM) central campus, located at -33.034796^o latitude and -71.595564^o longitude. All coordinates are measured using datum WGS84 as reference. These experimental locations included agricultural landscapes that together contained all the terrain types under evaluation. The trials were performed from 9 a.m. to 5 p.m. in summer and fall, obtaining shadowed and strongly illuminated scenes, as shown in Fig. 3.

The first stage of our experimental tests consisted in evaluating the classi-

fication performance as function of the variables given by the sensor. To this aim, a development dataset (acquired with the static setup) consisting of 520 and 300 images for training and testing, respectively was used. This stage also included a 10-fold cross-validation of the classification model obtained. Table 1 shows the performance of the proposed approach for different combinations of IR, color and depth streams, along with the features employed for each case. As can be noted, IR and color streams yielded the best results, whereas depth information did not contribute to improve the classification rates.

Table 1: Comparison of classification accuracy rates when employing different sensor streams to characterize the terrains.

Characteristic	Features	Accuracy (%)
Infra-red	Texture description	92.57
Colour	Texture description of grayscale image	81.20
Depth	Texture description	34.61
Infra-red + color	Texture description of IR image + color features	95.40
Infra-red + depth	Texture description of IR image + depth features	89.13
Color + depth	Texture description of grayscale image + depth features	80.61
Infra-red + color + depth	Texture description of IR image + color features + depth features	91.04

In addition, Figure 4 shows the confusion matrix for the best validation case. The far right column shows the accuracy for the output terrain types, whereas the row at the bottom shows the accuracy for each true class. Finally, the cell at the right bottom shows the overall accuracy of the classification. The most *conflictual* terrains were sand and pavement since some images from pavement were very similar to sand, specially from the color point of view. However, the general performance of the proposed approach was accurate enough to test it under real conditions.

Once the proposed methodology was validated, we ran a total of 17 trials navigating with the mobile robot through the three experimental landscapes. We acquired in average 1000 frames in each drive of the robot which were manually labelled to obtain the ground truth. In addition, and to provide

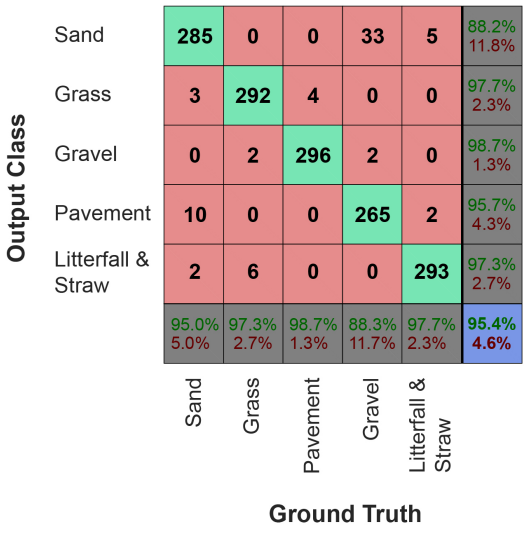


Figure 4: Confusion matrix for the best validation test. Only IR and color information from the sensing device was employed, yielding satisfactory detection rates. Two terrain types are often misclassified due to its visual similarity from the color and IR images.

an illustration of the data acquired by the sensor, Fig. 5 shows a 3D-color reconstruction of partial paths followed by the robot. Part of the measurements are noisy due to sunlight incidence and vibration of the vehicle, but in general, the paths are visually well rendered.

In order to evaluate the classification performance for this case, 5 statistics including accuracy (acc), precision (prec), recall (rec), specificity (spec) and F-score (Fs), were calculated. They were calculated as described by Fawcett (2006) and are aimed to measure, in a complementary way, the quality of the true positive and true negative detections, taking into account all the classification outputs. The results in terms of these metrics, along with the distance covered for each testing trial are summarized in Table 2. As can be noted, reduced performance is exhibited for some trials. A deeper analysis of the images in each one showed important variation in true positive and false negative detection rates for the reasons discussed in Section 5. However, in general, the proposed methodology to classify terrain types based only on their appearance, with data

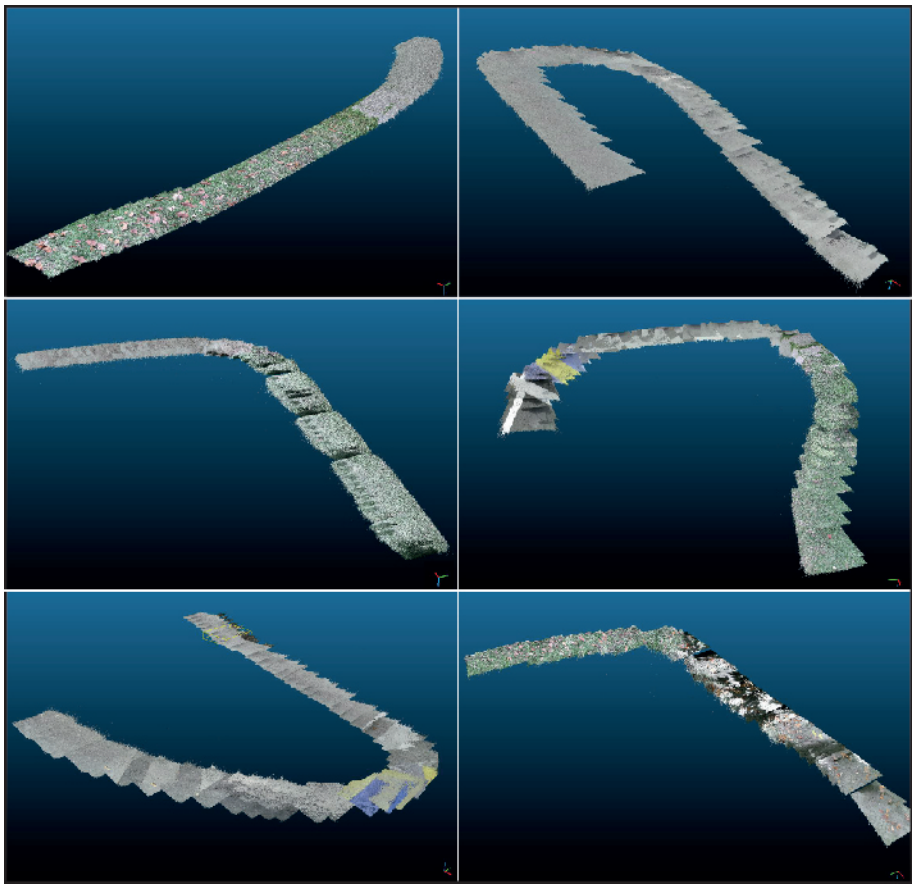


Figure 5: Three dimensional reconstructions of partial paths followed by the robot in different testing locations. The robot was manually driven on different roads in order to cover all the terrain classes classified in this work. In addition, the sensing device showed robustness when working outdoors, which allowed the scenes to be properly rendered.

acquired by a low cost sensor showed potential, even in presence of changing illumination and real driving conditions.

Table 2: Performance statistics for experiments conducted with the robot in real working conditions. A total of 17 trials were run within the three experimental locations, obtaining an overall of 15836 images.

Trial	Accuracy	Precision	Recall	Specificity	F-score	Distance(m)
1	0.80	0.80	0.80	0.95	0.80	144.63
2	0.76	0.76	0.76	0.92	0.76	121.46
3	0.77	0.77	0.77	0.94	0.77	87.63
4	0.80	0.80	0.80	0.93	0.80	173.57
5	0.91	0.91	0.91	0.98	0.91	325.70
6	0.92	0.92	0.92	0.98	0.92	124.93
7	0.72	0.72	0.72	0.91	0.72	98.69
8	0.86	0.86	0.86	0.97	0.86	119.3
9	0.92	0.92	0.92	0.98	0.92	132.31
10	0.96	0.96	0.96	0.98	0.96	107.80
11	0.79	0.79	0.79	0.95	0.79	163.00
12	0.87	0.87	0.87	0.97	0.87	168.22
13	0.82	0.82	0.82	0.95	0.82	136.11
14	0.98	0.98	0.98	0.98	0.98	286.40
15	0.73	0.73	0.73	0.91	0.73	146.46
16	0.87	0.87	0.87	0.96	0.87	424.95
17	0.82	0.82	0.82	0.95	0.82	477.05
mean	0.83	0.83	0.83	0.95	0.83	Total Distance: 3238.21 m

In addition to the previous results, Fig. 6 shows a confusion matrix that summarizes the outcomes obtained for all trials. In general, the proposed approach is able to achieve an accuracy of 83.00%, with reduced false positive and false negative detections for various terrain types in the field. However, the interclass similarity problem for pavement and sand observed in the validation stage is also exhibited in this case. Furthermore, sand and litterfall & straw also present this problem, which was not perceived in the validation stage. A close look into the images of these terrain types showed that some sandy locations were visually similar to litterfall & straw. Moreover, the most important reason for this outcome is the excessive black pixels with null information obtained in

numerous images, as will be discussed in Section 5. Despite of these issues, the overall performance of the proposed terrain classification system can be considered satisfactory, given the testing conditions and the low cost sensing hardware employed.

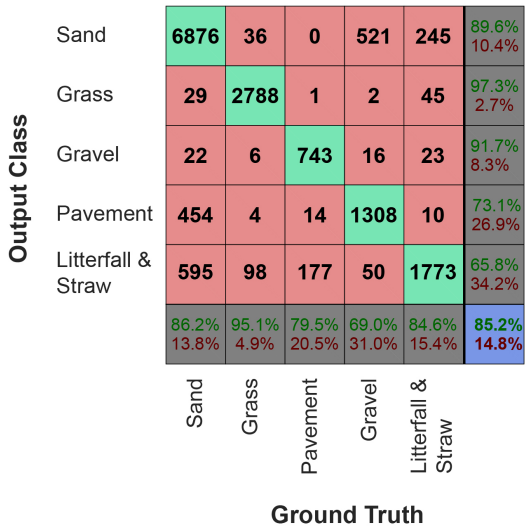
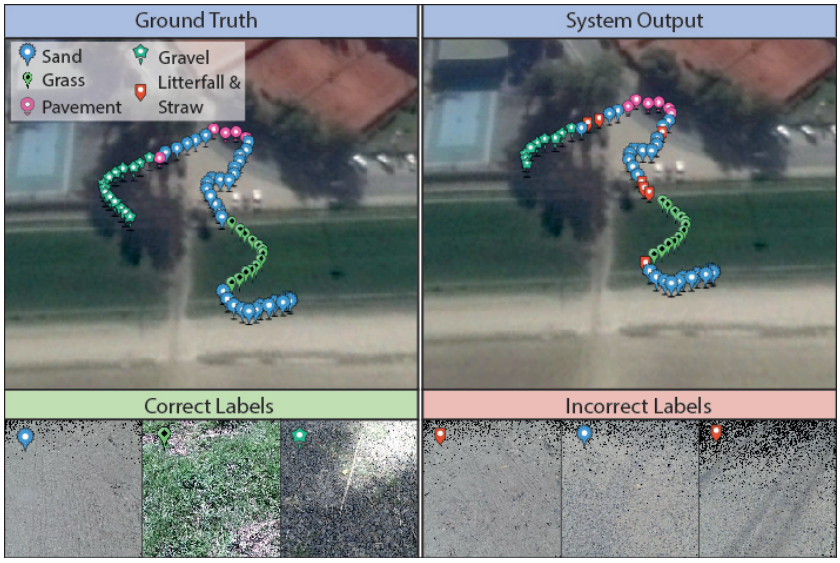


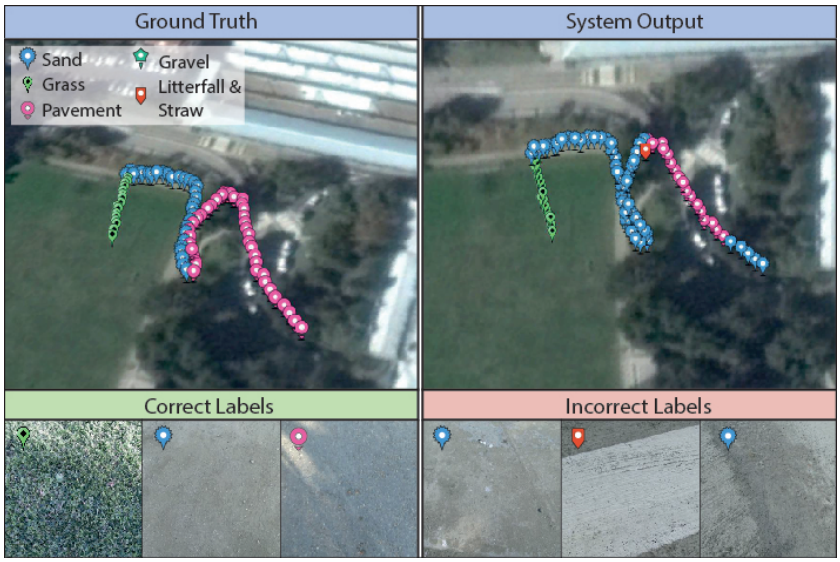
Figure 6: Confusion matrix that summarizes the overall performance of the proposed terrain classification methodology. It includes results of all tests (15836 images) conducted with the mobile robot through diverse agricultural scenes. Interclass visual similarity tends to decrease the classification performance, but in general, the terrains are correctly identified.

In order to illustrate the path followed by the robot, along with the labels obtained from the proposed approach, Fig 7a and Fig. 7b show a georeferenced and subsampled outcome for specific trials ran in two locations. They correspond to 65 labels from trials 6 and 15 which originally contained 647 and 870 frames, respectively.

All the experiments employed an on-board computer with a 2.20 Ghz Core i5 processor and 4 GB of RAM. With these characteristics, the processing time in average was 14.83 min per class for the training stage (with the extended dataset) and 0.51 s per image in the labelling stage. Since training the algorithm can be done offline before operating in field, and considering that the labelling



(a)



(b)

Figure 7: Georeferenced and subsampled outcome of the terrain classification system for two trials. The bottom row shows RGB images of the terrain captured by the sensing device, along with coloured markers representing the output of our classification system. The ground truth for these images are (from left to right): (a) sand, grass, gravel, sand, pavement, sand; and (b) grass, sand, pavement, pavement, pavement and pavement.

time is acceptable for vehicle moving at low speeds, the proposed approach is valid from the computational effort point of view. In addition, the memory required to store the learned model, as well as the texton dictionary is only 2 Mb, which would be important when employing low-performance computers with limited storage memory.

5. Discussion

Validation and real operation test results demonstrated the feasibility of the approach presented to classify different types of terrains. When comparing the suitability of the streams provided by the sensor, depth information tended to decrease the classification accuracy. According to the authors, this outcome is a result of two main reasons: the sensor resolution and the redundancy introduced due to the depth measuring methodology. The former causes terrains with slight changes in depth to look similar from the sensor point of view. The latter lies in the addition of extra processing of data from the same source, since depth is measured based on the IR reflectance received by the sensor. Thus, the best results in terms of classification accuracy were obtained when employing only the IR and color streams. With this regard, the confusion matrix depicted in Fig. 4, shows that sand and pavement classes are prone to missclassifications. This issue is due to the visual similarity between these two terrains which in some cases caused confusions even for the human labeller.

The visual interclass similarity issue identified in the validation stage was confirmed in the real testing stage. In this case, the problem was more evident and spread to other terrain types. For example, some sand images were confused with litterfall & straw since both were very similar from the IR texture and color point of view. Another minor issue that also contributed to reduce the true positive and true negative rates was related to the excessive artifacts in some images due to the robot motion and the presence of objects in the soil surface that modified the appearance of the captures. In order to illustrate these problems, Fig. 8 shows color images along with its IR frames of the two

issues previously described. The first and second columns of the image show pairs of highly similar color and IR images despite of their belonging to different classes. The third and fourth columns show the excessive artifacts introduced in some images that result in bad labelling. Most of these images in the field tests correspond to sandy terrain, that made the algorithm to classify it as litterfall & straw. This explains the outcome of the confusion matrix for this case (see Fig. 6), in which sand was missclassified as litterfall &straw 595 times.

In contrast to the previous concerns, most of the color and IR information acquired with the sensor allowed satisfactory results. Considering that we are employing only 520 training samples per class, the system showed robustness in different agricultural scenarios and testing conditions. Figure 9 shows various challenging captures that were correctly classified despite of having shadows and high illumination. It can be noted that color images are affected by these problems, but the IR streams are insensitive to them.

The points that we have identified to reduce the classification performance are related to the visual essence of our approach. Thus, improving the efficiency of the current approach seems suitable by using multi-sensor fusion. In this way, using the data acquired with an inertial sensor in a complementary way seems plausible. The output of our system can provide "a priori" information to be subsequently corroborated or corrected by the inertial measurements. In this case, the robustness under sunlight and the low cost of the proposed methodology could offer an increased versatility which would favour its commercial adoption.

Another important point of the presented work is the computational effort (0.51s for the labelling stage). For the trials with the robot operating at the reported linear and angular speeds, and taking into account the area in front of the robot covered by the sensor field of view, the processing time was satisfactory enough for real time operations. However, for speeds higher than those reported in this work, a lower level implementation of the proposed system is certainly required.

Finally, when comparing our approach with previous works, we observed

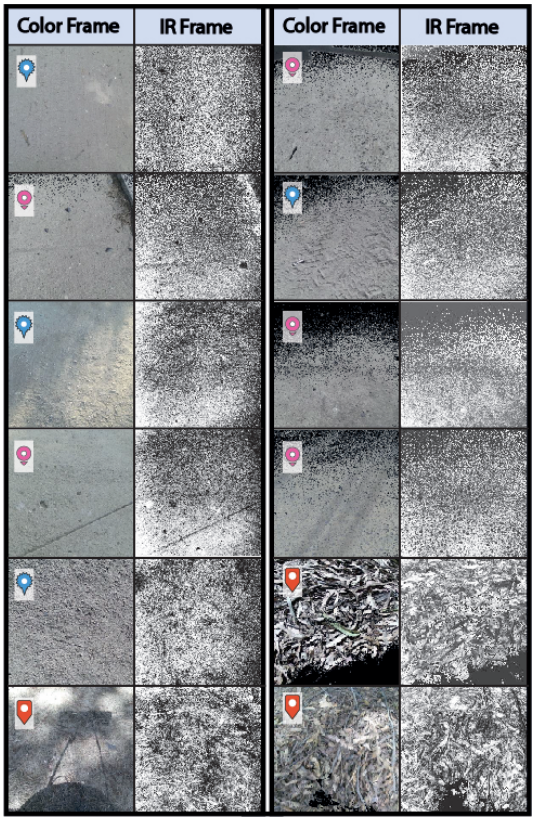


Figure 8: Examples of color images and their corresponding IR frames that show the visual similarity and excessive black pixels issues that resulted in missclassifications. The first and second columns show pairs of color and their IR images that illustrate the high similarity interclass for some captures. From top to bottom: sand, pavement, sand, pavement, sand and litterfall & straw. The third and fourth columns show pairs of color and their IR frames that illustrate the loss of information in some captures that also introduced error in the classification. The markers on each image show its true terrain type, according to the legend presented in Figure 7.

various pros and cons that are summarized in Table 3. Real time operation at low speeds, accuracy above 80% and low cost are the three strongest points that make the system presented in this brief comparable with those that used other sensors to the same aim: classify terrain types to analyse traversability.

Table 3: Comparison of the presented approach with others that employ different sensors and methodologies for terrain classification.

	Approach			
	Proposed Time of Flight IR description	Radar/ Laser	Stereo Vision	Self Learning
Pros	Requires a small dataset to be robust in changing illumination conditions.	Various works provide an statistical traversability analysis.	Can also include slip prediction.	Do not need large training datasets.
	Reduced cost.	Large field of view.	Large field of view.	Are verstaile when facing new terrains
	Overall accuracy higher than 85% using more than 15 000 images (more than 3.00 km of driving).	Robust in changing illumination conditions.	A geometric description of the terrain is available.	Some works provide a statistical traversability analysis
	A geometric description of the terrain can also be obtained.		Various terrains can be detected in the same image.	
Cons	Reduced field of view.	A finer classification in various terrain types can not be achieved.	Processing time is not suitable for online operation in some cases.	Rely on propioceptive information, which limits prediction capabilities.
	Only one label is obtained per image.	Cost is high for radar and 3D laser scanners.	Sensitive to changes in illumination.	Processing time is not suitable for online operation in some. cases
	Computation time needs to be improved.	Lack of color or spectral information.	Overall accuracy up to 76.4% using 1500 images for testing.	

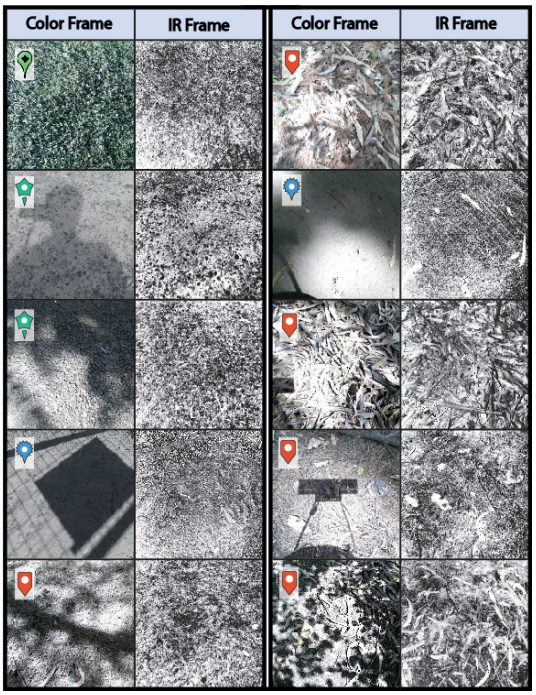


Figure 9: Examples of challenging color (1st and 3rd columns) and their IR streams (2nd and 4th columns) correctly classified by our system. These captures correspond to: grass, gravel, gravel, sand, litterfall & straw, litterfall & straw, sand, litterfall & straw, litterfall & straw and litterfall & straw, respectively. It can be noted that color frames are influenced by illumination, whereas IR frames are not, providing robustness for these conditions. The markers on each image show its true terrain type, according to the legend presented in Figure 7.

6. Conclusion and Further Work

This work presented the use of a visual approach to distinguish between agricultural terrains. Such information is intended to contribute to the traversability assessment of the surrounding terrain of agricultural machinery for supervision and inspection. Data acquired from a low cost sensor was used to characterize the predominant colors and texture of five representative agricultural terrain types including: sand, grass, pavement, gravel and litterfall & straw- covered. For this aim, histogram features were obtained based on exemplar local descriptors of 5×5 neighbourhoods per pixel (called textons), obtained by means of a

clustering algorithm.

The proposed system was first validated using a static experimental setup, that allowed to identify that only the color and IR streams provided enough information to obtain satisfactory results (accuracy higher than 95%). Subsequently, the system was tested in real working conditions by running several trials in agricultural landscapes. Mounting the sensor over the vehicle allowed to obtain an important view of its front region. However, the captured area was not enough to distinguish between representative patches of various terrain types in the same image.

The experimental results for the validation tests, as well as the driving experiments showed two main issues of the proposed approach. It was detected that acquisition glitches, along with high interclass similarity between sand and pavement reduced the accuracy and performance of our approach, specially when the sensor was mounted on the vehicle. However, in this conditions, our system is capable of obtaining an overall accuracy of 85.20%. Such outcome was considered satisfactory given the relative low number of training images and the changing illumination conditions observed during the robot navigation. Moreover, the processing time obtained in the method implementation admitted applications in real time at low speeds.

Given the previous discussion, the authors' further work will be focused on including other terrain types (e.g., rocky, or muddy), incorporating various sensors to enhance the classification, and implementing the proposed approach in a lower-level programming language. In addition, this work opened a new point of view to detect useful terrain characteristics for mobile robots navigation based on the spectral response (not only thermal) of the soil surface. Finally, a complete terrain characterization system would incorporate a terramechanical model of the ground-wheel interface to assess the traversability characteristics of the terrain in an integral manner.

Acknowledgements

This work was supported by the Chilean National Commission for Scientific and Technological Research under grant CONICYT- PCHA/Doctorado Nacional/2015-21150694, FONDECYT Grant 1171431, FB0008, DGIIP-UTFSM Chile and University of Lleida, Spain.

References

Al-Milli, S., Seneviratne, L.D., Althoefer, K., 2010. Track-terrain modelling and traversability prediction for tracked vehicles on soft terrain. Journal of Terramechanics 47, 151–160.

Alvarez, S., Vanrell, M., 2012. Texton theory revisited: A bag-of-words approach to combine textons. Pattern Recognition 45, 4315–4325. doi:10.1016/j.patcog.2012.04.032.

Andújar, D., Rueda-Ayala, V., Moreno, H., Rosell-Polo, J.R., Escolà, A., Valero, C., Gerhards, R., Fernández-Quintanilla, C., Dorado, J., Griepentrog, H.W., 2013. Discriminating crop, weeds and soil surface with a terrestrial LIDAR sensor. Sensors (Basel, Switzerland) 13, 14662–14675.

Angelova, A., Matthies, L., Helmick, D., Perona, P., 2007. Learning and Prediction of Slip from Visual Information. Journal of Field Robotics 24, 205–231.

Ball, D., Upcroft, B., Wyeth, G., Corke, P., English, A., Ross, P., Patten, T., Fitch, R., Sukkarieh, S., Bate, A., 2016. Vision-based Obstacle Detection and Navigation for an Agricultural Robot. Journal of Field Robotics 33, 1107–1130.

Botha, T.R., Els, P.S., 2015a. Digital image correlation techniques for measuring tyre road interface parameters: Part 2 Longitudinal tyre slip ratio measurement. Journal of Terramechanics 61, 101–112. doi:10.1016/j.jterra.2015.05.003.

Botha, T.R., Els, P.S., 2015b. Digital image correlation techniques for measuring tyre-road interface parameters: Part 1 Side-slip angle measurement on rough terrain. Journal of Terramechanics 61, 87–100. doi:10.1016/j.jterra.2015.04.004.

Brooks, C.A., Iagnemma, K., 2005. Vibration-based terrain classification for planetary exploration rovers. IEEE Transactions on Robotics 21, 1185–1190.

Brooks, C.A., Iagnemma, K., 2012. Self-Supervised Terrain Classification for Planetary Surface Exploration Rovers. Journal of Field Robotics 29, 445–468.

Broten, G., MacKay, D., Collier, J., 2012. Probabilistic obstacle detection using 2 1/2 D terrain maps, in: Proceedings of the 2012 9th Conference on Computer and Robot Vision, CRV 2012, pp. 17–23.

Falola, O., 2012. Drivable Region Detection for Autonomous Robots Applied to South African Underground Mining. Ph.D. thesis. University of CapeTown.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 861–874.

Fernandez, J.C., 2010. Characterization of Surface Roughness of Bare Agricultural Soils Using LIDAR. Ph.D. thesis. University of Florida.

Gao, H., Guo, J., Ding, L., Li, N., Liu, Z., Liu, G., Deng, Z., 2013. Longitudinal skid model for wheels of planetary exploration rovers based on terramechanics. Journal of Terramechanics 50, 327–343.

Häselich, M., Arends, M., Wojke, N., Neuhaus, F., Paulus, D., 2013. Probabilistic terrain classification in unstructured environments, in: Robotics and Autonomous Systems, Elsevier B.V.. pp. 1051–1059.

Ho, K., Peynot, T., Sukkarieh, S., 2013. A near-to-far non-parametric learning approach for estimating traversability in deformable terrain, in: IEEE International Conference on Intelligent Robots and Systems, pp. 2827–2833.

Hsu, C.W., Lin, C.J., 2002. A Comparison of Methods for Multiclass Support Vector Machines. IEEE Transactions on Neural Networks 13, 415–425.

Iagnemma, K., Ward, C.C., 2009. Classification-based wheel slip detection and detector fusion for mobile robots on outdoor terrain. Autonomous Robots 26, 33–46.

Ishigami, G., Kewlani, G., Iagnemma, K., 2009. Predictable mobility. IEEE Robotics and Automation Magazine 16, 61–70.

Lachat, E., Macher, H., Mittet, M.A., Landes, T., Grussenmeyer, P., 2015. First experiences with kinect V2 sensor for close range 3D modelling. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 40, 93–100.

Lau, D., 2013. The Science Behind Kinects or Kinect 1.0 versus 2.0. URL: http://www.gamasutra.com/blogs/DanielLau/20131127/205820/.

Li, S.Z., 2009. Markov Random Field Modeling in Image Analysis - Third Edition.

Manduchi, R., Castano, A., Talukder, A., Matthies, L., 2005. Obstacle Detection and Terrain Classification for Autonomous Off-Road Navigation. Autonomous Robots 18, 81–102.

Marinello, F., Pezzuolo, A., Gasparini, F., Arvidsson, J., Sartori, L., 2015. Application of the Kinect sensor for dynamic soil surface characterization. Precision Agriculture 16, 601–612.

Michel, D.C., 2012. System of Terrain Analysis , Energy Estimation and Path Planning for Planetary Exploration by Robot Teams. Ph.D. thesis. University of Western Ontario.

Ojeda, L., Borenstein, J., Witus, G., Karlsen, R., 2006. Terrain Characterization and Classification with a Mobile Robot. Journal of Field Robotics 23, 103–122.

Ono, M., Fuchs, T.J., Steffy, A., Maimone, M., Yen, J., 2015. Risk-aware Planetary Rover Operation : Autonomous Terrain Classification and Path Planning, in: Aerospace Conference, pp. 1–10.

Otsu, K., Ono, M., Fuchs, T.J., Baldwin, I., Kubota, T., 2016. Autonomous Terrain Classification With Co- and Self-Training Approach. IEEE Robotics and Automation Letters 1, 814–819.

Park, B., Kim, J., Lee, J., 2012. Terrain Feature Extraction and Classification for Mobile Robots Utilizing Contact Sensors on Rough Terrain, in: International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012), pp. 846–853.

Prado, A., Auat Cheein, F.A., Torres-Torriti, M., 2016. Probabilistic Approaches for Self-tuning Path Tracking Controllers using Prior Knowledge of the Terrain, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3095–3100.

Reina, G., Milella, A., Rouveure, R., Nielsen, M., Worst, R., Blas, M.R., 2016. Ambient awareness for agricultural robotic vehicles. Biosystems Engineering 146, 114–132.

Reina, G., Milella, A., Underwood, J., 2012. Self-learning classification of radar features for scene understanding. Robotics and Autonomous Systems 60, 1377–1388.

Rosell-Polo, J.R., Auat Cheein, F., Gregorio, E., Andújar, D., Puigdomènech, L., Masip, J., Escolà, A., 2015. Advances in Structured Light Sensors Applications in Precision Agriculture and Livestock Farming. Advances in Agronomy 133, 71–112.

Schwiegerling, J., 2004. Field Guide to Visual and Ophthalmic Optics.

Stolpe, N., 2002. Clasificaciones Interpretativas. Technical Report. National Institute of Agricultural Research of Chile. URL: `http://www2.inia.cl/medios/biblioteca/serieactas/NR29051.pdf`.

584 Taheri, S., Sandu, C., Taheri, S., Pinto, E., Gorsich, D., 2015. A technical survey
585 on Terramechanics models for tire-terrain interaction used in modeling and
586 simulation of wheeled vehicles. Journal of Terramechanics 57, 1–22.

587 Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel,
588 J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C.,
589 Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek,
590 L.e., Koelen, C., Markey, C., Rummel, C., Niekerk, J.V., Jensen, E., Alessan-
591 drini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A., Ma-
592 honey, P., 2006. Stanley : The Robot that Won the DARPA Grand Challenge.
593 Journal of Field Robotics 23, 661–692.

594 Varma, M., Varma, M., Zisserman, A., Zisserman, A., 2009. A Statistical Ap-
595 proach To Material Classification Using Image Patch Exemplars. IEEE Tran-
596 scations on Pattern Analysis and Machine Intelligence 31, 2032–2047.

597 Varma, M., Zisserman, A., 2002. Classifying Images of Materials : Achieving
598 Viewpoint and Illumination Independence. Lecture notes in computer science
599 , 255–271.

600 Varma, M., Zisserman, A., 2005. A statistical approach to texture classification
601 from single images. International Journal of Computer Vision 62, 61–81.

602 Woods, M., Guivant, J., Katupitiya, J., 2015. Terrain Classification and Seg-
603 mentation using Non-Semantic Range Data, in: Australasian Conference on
604 Robotics and Automation (ACRA), Canberra.

605 Xia, C., Wang, L., Chung, B.K., Lee, J.M., 2015. In Situ 3D Segmentation of
606 Individual Plant Leaves Using a RGB-D Camera for Agricultural Automation.
607 Sensors (Basel, Switzerland) 15, 20463–79.

608 Xu, Y., Huang, S., Ji, H., Fermüller, C., 2012. Scale-space texture description
609 on SIFT-like textons. Computer Vision and Image Understanding 116, 999–
610 1013. doi:10.1016/j.cviu.2012.05.003.

Xue, J., Zhang, L., Grift, T.E., 2012. Variable field-of-view machine vision based row guidance of an agricultural robot. Computers and Electronics in Agriculture 84, 85–91.

Yandun, F., Gregorio, E., Zuñiga, M., Escolà, A., Rosell-Polo, J.R., Auat Cheein, F., 2016. Classifying Agricultural Terrain for Machinery Traversability Purposes, in: AGRICONTROL 2016: The 5th IFAC Conference on Sensing, Control and Automation for Agriculture, Elsevier. pp. 457–462.

Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C., 2007. Local features and kernels for classification of texture and object categories: A comprehensive study. International Journal of Computer Vision 73, 213–238. doi:10.1007/s11263-006-9794-4, arXiv:arXiv:1011.1669v3.

Zou, Y., Chen, W., Xie, L., Wu, X., 2014. Comparison of different approaches to visual terrain classification for outdoor mobile robots. Pattern Recognition Letters 38, 54–62.